



King's Research Portal

DOI:

[10.1175/JHM-D-12-0132.1](https://doi.org/10.1175/JHM-D-12-0132.1)

Document Version

Publisher's PDF, also known as Version of record

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Peña-Arancibia, J. L., Van Dijk, A. I. J. M., Renzullo, L. J., & Mulligan, M. (2013). Evaluation of precipitation estimation accuracy in reanalyses, satellite products, and an ensemble method for regions in Australia and south and east Asia. *JOURNAL OF HYDROMETEOROLOGY*, 14(4), 1323-1333. <https://doi.org/10.1175/JHM-D-12-0132.1>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Evaluation of Precipitation Estimation Accuracy in Reanalyses, Satellite Products, and an Ensemble Method for Regions in Australia and South and East Asia

JORGE L. PEÑA-ARANCIBIA

CSIRO Land and Water, Canberra, Australian Capital Territory, Australia, and Faculty of Earth and Life Sciences, VU University Amsterdam, Amsterdam, Netherlands, and Environmental Monitoring and Modelling Research Group, Department of Geography, King's College London, London, United Kingdom

ALBERT I. J. M. VAN DIJK* AND LUIGI J. RENZULLO

CSIRO Land and Water, Canberra, Australian Capital Territory, Australia

MARK MULLIGAN

Environmental Monitoring and Modelling Research Group, Department of Geography, King's College London, London, United Kingdom

(Manuscript received 6 September 2012, in final form 10 December 2012)

ABSTRACT

Precipitation estimates from reanalyses and satellite observations are routinely used in hydrologic applications, but their accuracy is seldom systematically evaluated. This study used high-resolution gauge-only daily precipitation analyses for Australia (SILO) and South and East Asia [Asian Precipitation—Highly-Resolved Observational Data Integration Towards Evaluation (APHRODITE)] to calculate the daily detection and accuracy metrics for three reanalyses [ECMWF Re-Analysis Interim (ERA-Interim), Japanese 25-yr Reanalysis (JRA-25), and NCEP–Department of Energy (DOE) Global Reanalysis 2] and three satellite-based precipitation products [Tropical Rainfall Measuring Mission (TRMM) 3B42V6, Climate Prediction Center morphing technique (CMORPH), and Precipitation Estimation from Remotely Sensed Imagery Using Artificial Neural Networks (PERSIANN)]. A depth-frequency-adjusted ensemble mean of the reanalyses and satellite products was also evaluated. Reanalyses precipitation from ERA-Interim in southern Australia (SAu) and northern Australasia (NAu) showed higher detection performance. JRA-25 had a better performance in South and East Asia (SEA) except for the monsoon period, in which satellite estimates from TRMM and CMORPH outperformed the reanalyses. In terms of accuracy metrics (correlation coefficient, root-mean-square difference, and a precipitation intensity proxy, which is the ratio of monthly precipitation amount to total days with precipitation) and over the three subdomains, the depth-frequency-adjusted ensemble mean generally outperformed or was nearly as good as any of the single members. The results of the ensemble show that additional information is captured from the different precipitation products. This finding suggests that, depending on precipitation regime and location, combining (re)analysis and satellite products can lead to better precipitation estimates and, thus, more accurate hydrological applications than selecting any single product.

* Current affiliation: Fenner School of Environment and Society, College of Medicine, Biology and Environment, Australian National University, Canberra, ACT, Australia.

Corresponding author address: Jorge L. Peña-Arancibia, CSIRO Land and Water, GPO 1666, Black Mountain, Canberra ACT 2601, Australia.
E-mail: jorge.penaarancibia@csiro.au

1. Introduction

The accuracy of precipitation estimates to a great extent determines the accuracy of hydrological model outputs (Fekete et al. 2004; Fernandes et al. 2008; Voisin et al. 2008; Pan et al. 2010; Getirana et al. 2011; Van Dijk and Renzullo 2011; Yong et al. 2012). Gridded precipitation analysis based on gauging can be of dubious

quality in areas where gauge or radar networks do not exist or are sparse, for example, in much of the tropics. Several precipitation estimates derived from satellite data or modeled through retrospective weather forecast model analysis (reanalysis) provide estimates that are independent from gauge networks. Both types of precipitation estimates have been increasingly used in hydrological applications [e.g., for reanalysis (Dedong et al. 2007; Li et al. 2009; Yan et al. 2010; Miguez-Macho and Fan 2012) and for satellite (Shrestha et al. 2008; Behrangi et al. 2011; Khan et al. 2012; among many others)].

Previous studies evaluating reanalyses and satellite precipitation estimates in areas with dense gauge or radar coverage suggest that convective precipitation (more typical of warmer seasons and lower latitudes) is better characterized by satellite precipitation, whereas frontal system precipitation (more typical of cooler seasons and higher latitudes) is better characterized by reanalysis (e.g., Gottschalk et al. 2005; Ebert et al. 2007; Ruane and Roads 2007; Tian et al. 2009; Sapiano and Arkin 2009; Vila et al. 2010). Estimates from these products can be very different, particularly over tropical areas with high precipitation (Bosilovich et al. 2008; Tian and Peters-Lidard 2010). The incorporation of rain gauge data to correct magnitudes and frequencies can reduce total errors and bring the intensity distribution for heavy precipitation closer to the gauge data (Ebert et al. 2007). It is also noted that more recent reanalyses have improved precipitation estimates for tropical areas, although notable biases still exist (Betts et al. 2006, 2009; Bosilovich et al. 2008; Uppala et al. 2007).

The above summary suggests that reanalysis and satellite datasets can be complementary. This would be particularly relevant in areas where adjustments are difficult or impossible because of the scarcity of rain gauge networks. The aims of this paper are to 1) evaluate and compare daily satellite and reanalysis precipitation estimates routinely used in large-scale hydrologic model applications against precipitation analysis based on dense ground networks in Australia and South and East Asia and 2) evaluate and compare a depth-frequency-adjusted ensemble mean of the products (see definition in section 2 below). The performance metrics are chosen to establish which precipitation product performs best for detection (occurrence) and estimation accuracy for daily precipitation (i.e., how close to the observed magnitude and/or frequency) in three subdomains with different precipitation regimes. Section 2 introduces the reanalyses, satellite, and evaluation precipitation datasets and performance metrics. Section 3 presents results of the performance evaluation experiments. Section 4 discusses the results and draws conclusions.

2. Data and methodology

Three recent reanalysis precipitation datasets with global coverage are considered in this paper: 1) the National Centers for Environmental Prediction–Department of Energy Global Reanalysis 2 (NCEP–DOE; Kanamitsu et al. 2002), 2) the European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis Interim (ERA-Interim; Dee et al. 2011), and 3) the Japanese 25-yr Reanalysis (JRA-25; Onogi et al. 2007). These reanalyses build and improve on earlier reanalysis versions by improving the forecasting model physics and incorporating new satellite and other conventional data.

Also included are three quasi-global satellite-based precipitation products that combine multiple microwave and infrared sensors: 1) the bias-corrected Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (Huffman et al. 2007) 3B42V6, which uses monthly gauge observations to scale precipitation estimates; 2) the Climate Prediction Center (CPC) morphing technique (CMORPH; Joyce et al. 2004); and 3) the Precipitation Estimation from Remotely Sensed Imagery using Artificial Neural Networks (PERSIANN; Sorooshian et al. 2000).

An ensemble of the six products was derived by calculating the simple mean daily precipitation and adjusting it to the depth-frequency distribution function of gauge-only daily precipitation analyses (used as evaluation data) by mapping the full spatiotemporal distribution of the ensemble estimates to that of the gauge analysis. In other words, if $\text{prob}^* = \text{prob}(P_{\text{ens},i})$ is the probability of the ensemble mean precipitation P_{ens} on day i , then the depth-frequency-adjusted precipitation estimate P_{adj} is $P_{\text{adj}} = P_{\text{obs}}(\text{prob}^*)$, with P_{obs} being the gauge analysis time series used. The adjustment is performed using all daily data for the grid cells selected from the gauge analysis products; consequently, the ensemble is not fully independent of the evaluation data.

The evaluation is performed at a daily temporal scale using two high-resolution gauge-only daily precipitation analyses available in Australia (SILO; Jeffrey et al. 2001) and South and East Asia [Asian Precipitation—Highly-Resolved Observational Data Integration Towards Evaluation (APHRODITE; Yatagai et al. 2012)]. All data were resampled to 1° resolution (simple averaging) as a compromise between the spatial resolutions of the different products (satellite estimate resolution data were 0.25° , whereas reanalyses ranged from 0.7° to 2.5°). Only grid cells with a density of more than one gauge per 500 km^2 were considered (see Fig. 1 for location of the grid cells). The common period for all data was 2003–07, and the time series of each precipitation product had less than 5% of days with no data. A threshold of 1 mm day^{-1}

was used to discriminate between “rain” and “no rain” in order to eliminate very light intensity “drizzle” that does not significantly contribute to daily precipitation but could have an undue impact on detection metrics. To account for differences in precipitation regime, the geographical domain was divided into three regions: southern Australia (SAu), mostly dominated by synoptic system precipitation during austral winter; northern Australasia (NAu), mostly dominated by convective precipitation during summer; and South and East Asia (SEA), mostly dominated by monsoon precipitation.

First, precipitation bias error estimates on annual and monthly time scales are computed following Adler et al. (2012). The standard deviation σ of the six products is used as a measure of the bias error. The dispersion among the product estimates captured in σ showcases the different physical assumptions and nature of both satellite and reanalyses precipitation retrievals. Subsequently, detection and accuracy metrics were computed for each grid cell. Every day in the estimated and gauge analysis was classified following Ebert et al. (2007) as a hit (H , observed precipitation correctly detected), miss (M , observed precipitation not detected by product), or false alarm (F , precipitation detected but none observed). The probability of detection, $POD = H/(H + M)$, gives the fraction of precipitation occurrences correctly detected (range 0–1 and a perfect score of 1). The false alarm ratio, $FAR = F/(H + F)$, gives the wrongly detected precipitation (range 0–1 and a perfect score of 0). The frequency bias, $FB = (H + F)/(H + M)$, gives the ratio of the estimated to observed precipitation frequency (range 0– ∞ and a perfect score of 1). The equitable threat score (ETS), used as an overall performance metric, gives the fraction of precipitation that was correctly detected, adjusted for correct detections (H_e) that would be expected because of random chance: $ETS = (H - H_e)/(H + M + F - H_e)$, where $H_e = (H + M)(H + F)/N$ and N is the total number of estimates (range $-1/3$ –1, a perfect score of 1 and 0 indicating no skill).

Accuracy metrics used were correlation r , root-mean-square difference (RMSD), and a precipitation intensity proxy, namely, the percentage difference of the ratio of monthly precipitation amount to the total number of days with precipitation (MPDR). Both detection and accuracy metrics were mapped for spatial patterns and examined. Results were also stratified by season to assist in interpretation. Finally, monthly and subdomain aggregated time series were plotted to detect any evidence for drifts or step changes.

3. Results

The mean annual precipitation of the six precipitation products used here is shown in Fig. 1a. The measure of

the bias error, the mean annual precipitation standard deviation σ of the products, is shown in Fig. 1b. As expected, higher σ values occur in grid cells with higher precipitation, with the highest values ($>1500 \text{ mm yr}^{-1}$) occurring in grid cells located in the intertropical convergence zone (ITCZ), particularly in insular Southeast Asia. The mean annual σ in SEA is 554 mm yr^{-1} , whereas it is 324 and 170 mm yr^{-1} in NAu and SAu, respectively. The months of January and July are used as an example of monthly bias. The ITCZ moves southward, and during January grid cells in northern Australia have the highest precipitation ($>200 \text{ mm month}^{-1}$) and σ ($>100 \text{ mm month}^{-1}$) (Figs. 1c,d). The mean January σ in SEA is 26 mm yr^{-1} , whereas it is 76 and 15 mm yr^{-1} in NAu and SAu, respectively. In July, the ITCZ shifts northward, and many grid cells in SEA are affected by the Asia–Pacific monsoon, with higher July precipitation occurring in grid cells in Japan, Nepal, southern China, and Southeast Asia ($>400 \text{ mm month}^{-1}$) (Fig. 1e). Higher July σ ($>100 \text{ mm month}^{-1}$) is observed not only in these grid cells, but also in southwest Australia and Tasmania (Fig. 1f).

Figure 2 shows percentage frequency of exceedance curves for the six products (for daily precipitation $>1 \text{ mm}$), the simple ensemble mean, and the depth-frequency-adjusted ensemble (data are aggregated over the whole geographical domain). All satellite products have lower frequencies than the reference for mean precipitation depths $<10 \text{ mm day}^{-1}$, whereas reanalyses agree reasonably well (Fig. 2a). The exceptions are NCEP–DOE, which exceeds reference depths almost across the range, and the simple ensemble mean, in which the simple averaging of all products enhances light precipitation depths (Fig. 2a). Conversely, only the bias-corrected TRMM 3B42V6 and the depth-frequency-adjusted ensemble show good agreement for mean precipitation depths $>50 \text{ mm day}^{-1}$ (Fig. 2b). Not surprisingly, the depth-frequency-adjusted ensemble shows this good agreement across the whole precipitation depth range.

In terms of ETS computed for the full time series, ERA-Interim performed best in SAu and parts of NAu close to SAu (Fig. 3a). NCEP–DOE performed best in parts of western and southern Australia, and JRA-25 performed best in most of Japan and South Korea. CMORPH and TRMM performed best in Southeast Asia. Results in continental Asia were mixed, with satellite products and the ensemble performing best in the tropics and reanalyses performing best in midlatitudes. From June to August (JJA), satellite data and the ensemble performed best in most of continental and southeast Asia and in areas in Japan most affected by the monsoon (Fig. 3b). JRA-25 and the ensemble generally performed best in December–February (DJF), except in

Precipitation

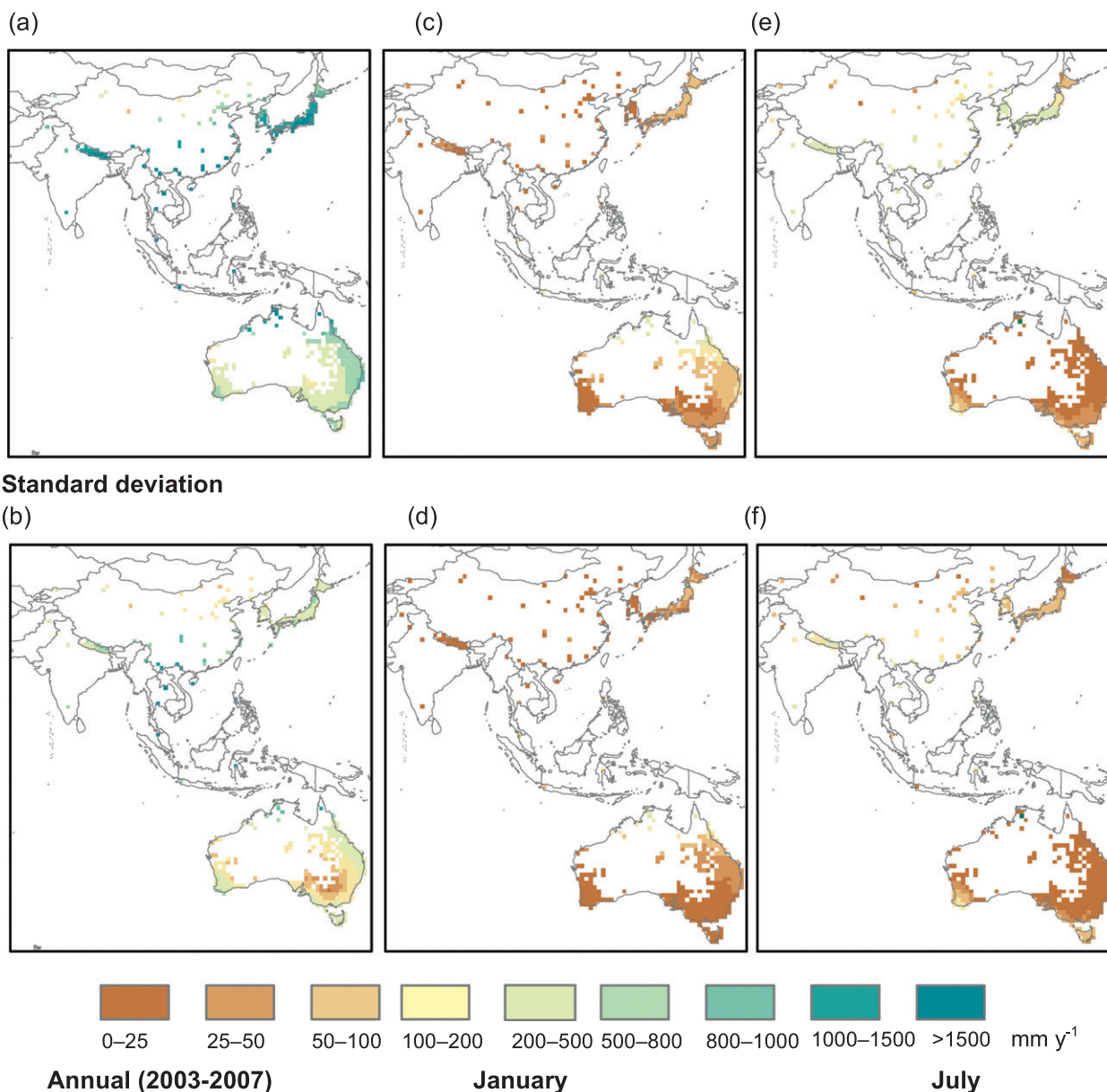


FIG. 1. (a) Mean precipitation and (b) standard deviation for the ensemble of the six precipitation products used in this study (not including the ensemble) for the years 2003–07. (c),(d) As in (a),(b) but for all January months. (e),(f) As in (a),(b), but for all July months.

insular Southeast Asia, where satellite products outperformed reanalyses and the ensemble (Fig. 3c).

Box plots in Fig. 4 highlight the superior detection performance of reanalyses for all geographical subdomains, with the exception of JJA (monsoon) in SEA (Figs. 4a–c). ERA-Interim performs better than satellite data in NAu during DJF. The ensemble shows performance somewhat intermediate to both product types. Seasonal variation in performance was not observed in SAu, but

there was an improvement in CMORPH and TRMM ETS during DJF in NAu (Fig. 4b). In SEA, ETS for ERA-Interim and JRA-25 were higher than satellite, except for JJA, where CMORPH and TRMM were better (Fig. 4a).

In terms of accuracy metrics, the spatial results did not show clear seasonal variations; thus, results are presented for all months combined. For r , JRA-25 performed best in most of southeast and parts of southwest Australia,

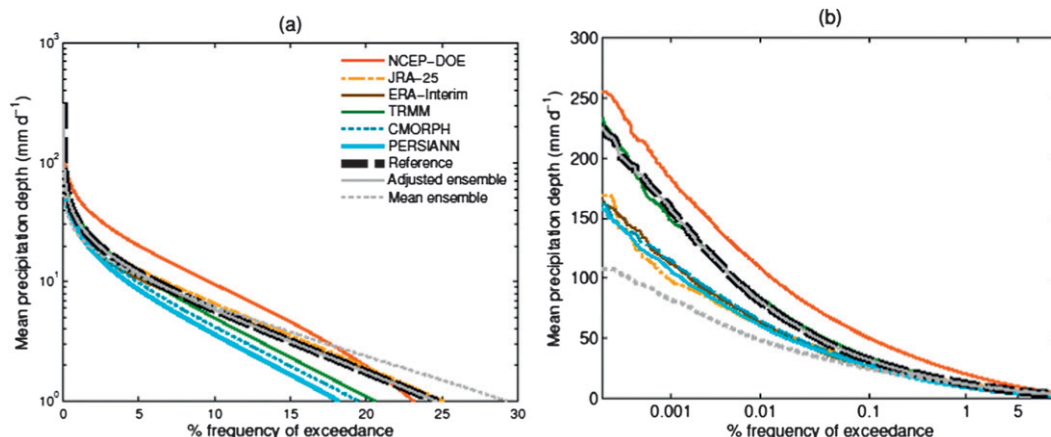


FIG. 2. (a) Percentage frequency of exceedance curves for daily precipitation (>1 mm) aggregated over the whole geographical domain. (b) As in (a), but the x axis is zoomed-in for higher mean precipitation depths.

whereas NCEP-DOE did so in parts of southern Australia, Tasmania, and southwest Australia (Fig. 5a).

ERA-Interim performed best in north Australia, whereas a combination of satellite and the ensemble performed best in the tropics. Box plots for all months show that r for ERA-Interim in both SAu and NAu were better than for satellite precipitation, with higher r for NAu (Fig. 6a). The ensemble performed best in Nepal, close to the coastline in China, and in part of Japan. JRA-25 and ERA-Interim had better performance in inland north China and also in some parts of Japan. For SEA, mean r is substantially higher (0.62) than in SAu (0.17) and NAu (0.35), with TRMM being superior and CMORPH comparable to JRA-25 and ERA-Interim

(Fig. 6a). The ensemble outperformed the other products in all subdomains.

ERA-Interim had the lowest RMSD in most of SAu and NAu (Fig. 5b). In China, satellite data generally performed best close to the coastline and reanalyses in the north, the ensemble in Nepal, and JRA-25 in most of Japan and South Korea. Box plots show that RMSD was slightly lower for ERA-Interim and JRA-25 in all subdomains. Errors in NCEP-DOE were systematically higher than the other datasets, whereas the ensemble RMSD was comparable to the best results in all subdomains (Fig. 6b).

MPDR results in SAu and NAu were mixed, but overall, the best performer was JRA-25, followed by the

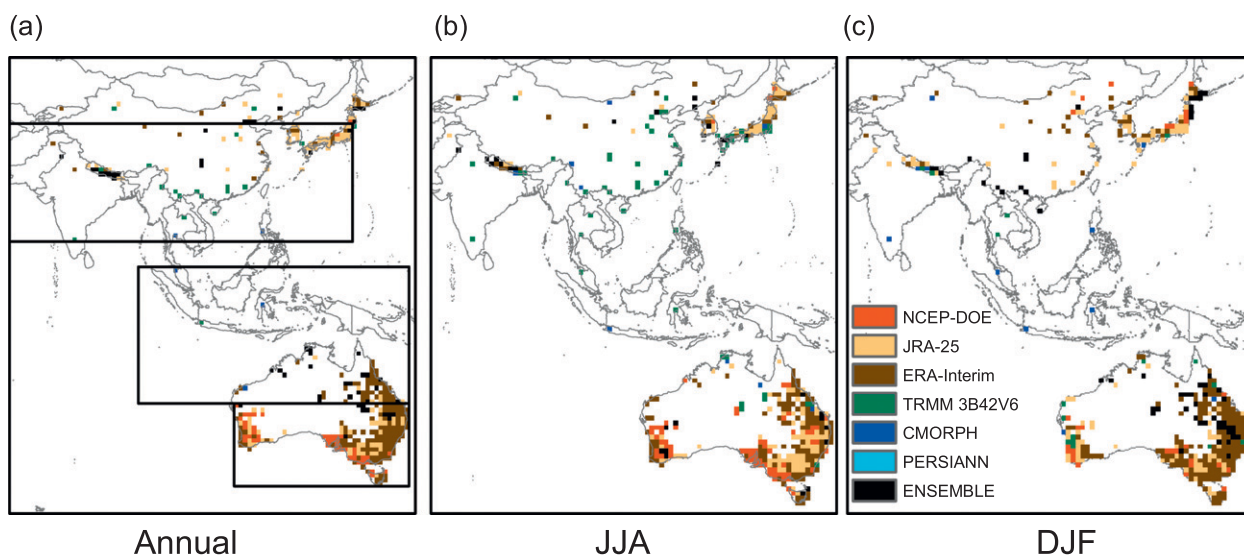


FIG. 3. Best performing product for estimating the occurrence of daily precipitation (>1 mm) in terms of the ETS for 2003–07 in each grid cell: (a) for all months, (b) for JJA, and (c) for DJF. Rectangles in (a) define the geographical extent for SAu, NAu, and SEA.

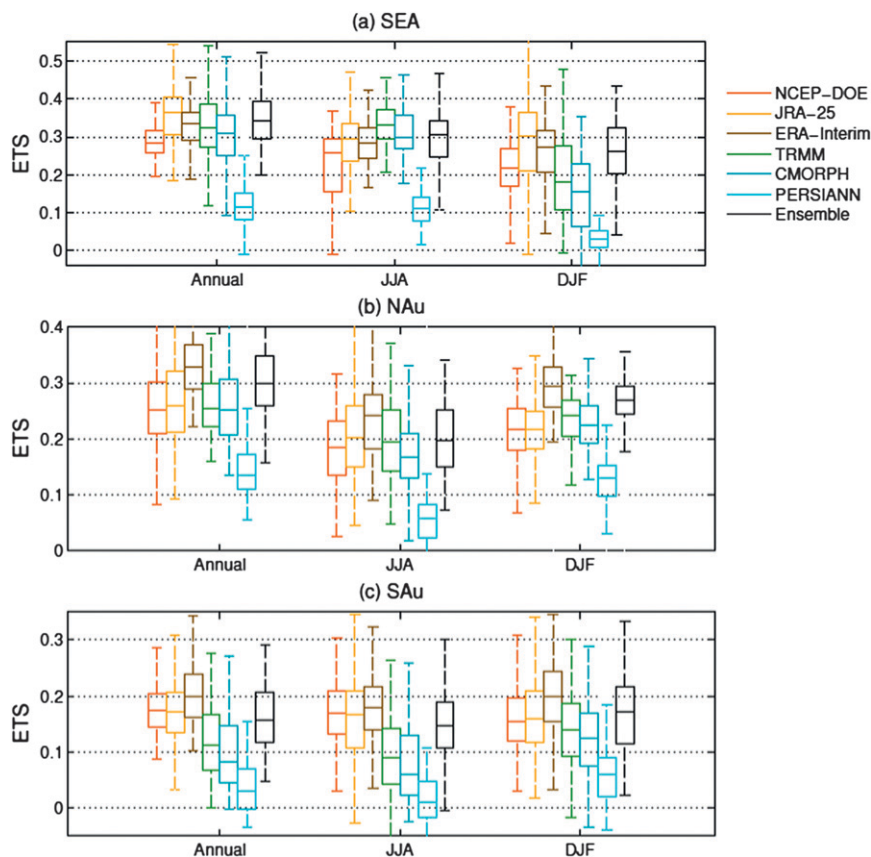


FIG. 4. Box plots showing the performance of the estimated occurrence of daily precipitation (>1 mm) in terms of ETS aggregated over three geographical subdomains for all months, JJA, and DJF: (a) SEA, (b) NAU, and (c) SAU. Tops and bottoms of each box are the 25th and 75th percentiles and whiskers are the 5th and 95th percentiles, respectively.

ensemble and ERA-Interim (Figs. 5c, 6c). Results were mixed in Japan. In China, satellite products and the ensemble generally performed best inland, whereas JRA-25 did so close to the coastline (Fig. 5c). Generally (besides NCEP-DOE), all products had less than 20% difference with observed MPDR with the exception of CMORPH and PERSIANN in SEA, which underestimated MDPR by 22% and 23%, respectively (Fig. 6c). ERA-Interim and PERSIANN systematically underestimated and TRMM and NCEP-DOE systematically overestimated MPDR (Fig. 6c).

Time series of monthly averaged ETS over the whole domain showed some seasonal variation, with an increase roughly during JJA and DJF, and a step change to reduced ETS for PERSIANN precipitation after 2005 (Fig. 7a). No clear patterns are evident for reanalysis data. The same step change in PERSIANN is present in the r time series, with again no obvious patterns for the other precipitation datasets (Fig. 7b). An analysis of PERSIANN FAR and POD over the subdomains

revealed that an increase in false detections in SAU and a decrease in correct detections in SAU and SEA were the cause for the step change (not shown). This likely affected r , but only for small precipitation depths, as RMSD and MPDR appear not much affected. For all products, RMSD time series showed an increase in errors during JJA and less so during DJF, with NCEP-DOE having the largest errors (Fig. 7c). NCEP-DOE and, surprisingly, the bias-corrected TRMM produced high positive MPDR values through the analysis period; CMORPH mostly produced positive values, and the rest of the datasets mostly produced low negative MPDR values.

Table 1 shows the product ranking for detection and accuracy metrics over the whole geographical domain and for all months. The depth-frequency-adjusted ensemble mean outperformed both satellite and reanalyses for most metrics. Among individual products, JRA-25 outperformed the others in most metrics, but its high FB suggested that it tends to over predict precipitation

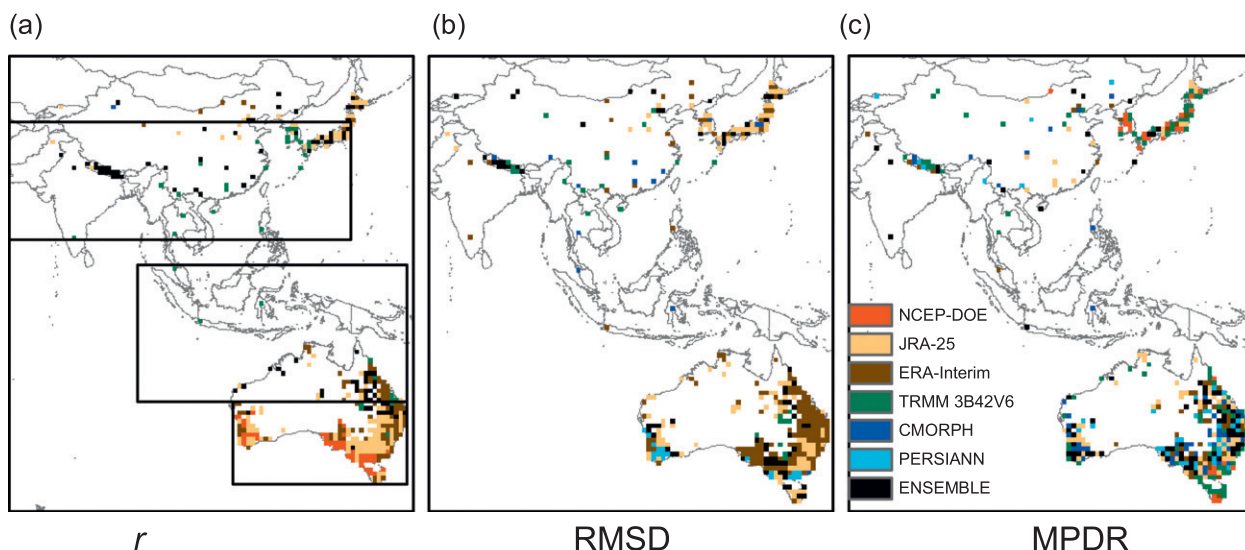


FIG. 5. Best performing product for accuracy statistics of daily precipitation (>1 mm) for 2003–07 in each grid cell: (a) correlation, (b) RMSD (mm day^{-1}), and (c) MPDR (%).

occurrence. CMORPH agreed better with observed MPDR; however, this is possibly because of compensating underprediction in SEA and overprediction in SAu and NAu (Fig. 6c).

4. Discussion and conclusions

Three reanalyses (ERA-Interim, JRA-25, and NCEP–DOE) and three satellite-based precipitation products

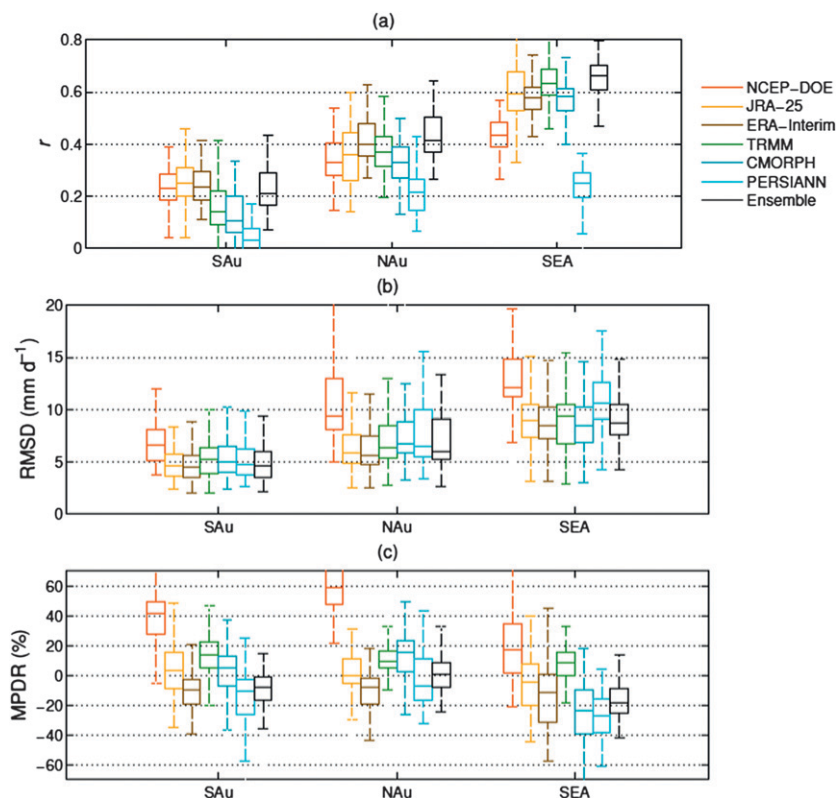


FIG. 6. Box plots showing accuracy statistics of daily precipitation (>1 mm) aggregated over three geographical subdomains: (a) r , (b) RMSD, and (c) MPDR. Tops and bottoms of each box are the 25th and 75th percentiles and whiskers are the 5th and 95th percentiles, respectively.

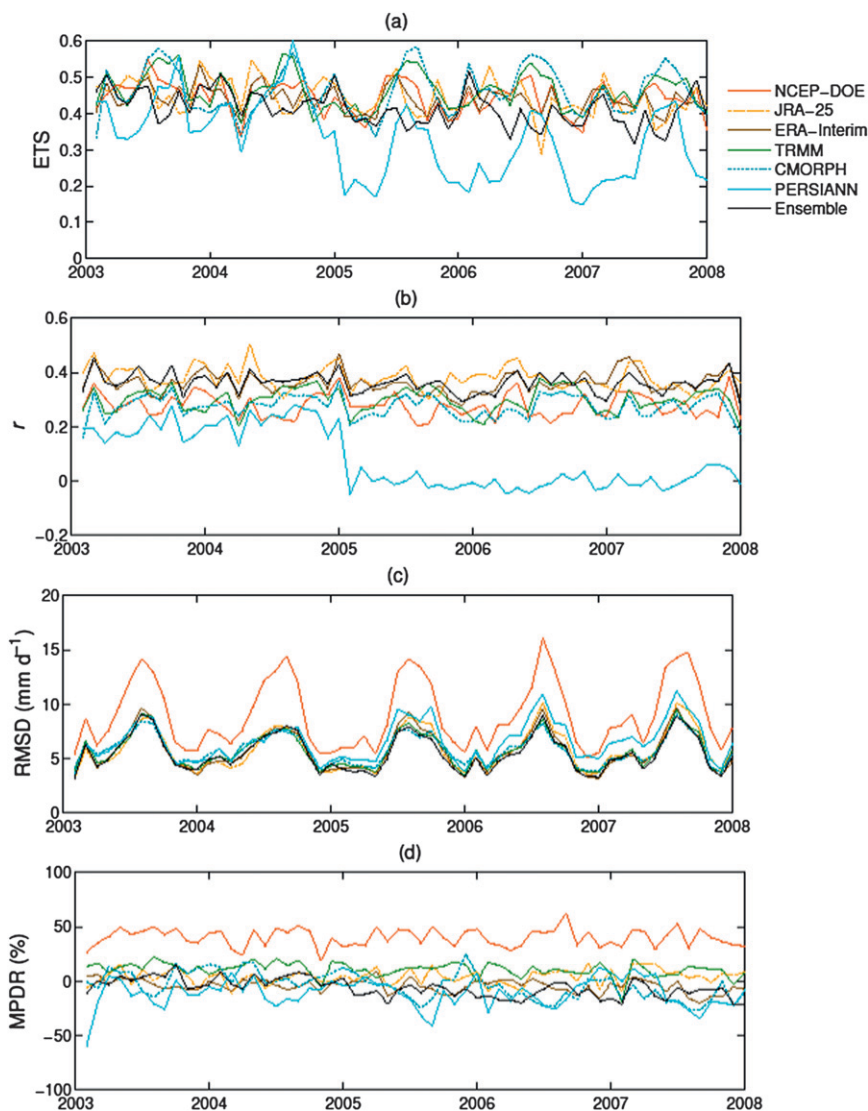


FIG. 7. Monthly time series for 2003–07 of (a) ETS, (b) r , (c) RMSD, and (d) MPDR averaged over the entire geographical domain.

(TRMM 3B42V6, CMORPH, and PERSIANN) were systematically evaluated, along with a depth-frequency-adjusted ensemble of the products, against analysis data in relatively well gauged areas in Australia and South and East Asia. Large bias errors (in terms of standard deviation of the products) indicated areas in which choice of precipitation estimates used in hydrologic applications should be carefully considered. Bias errors were large in some areas of high precipitation, such as the ITCZ, and also in high latitudes during winter months (southern Australia and Tasmania).

Analysis of precipitation ETS showed that reanalyses generally outperformed satellite precipitation estimates in all subdomains, except for JJA in SEA, that is, the months affected by the Asia–Pacific monsoon (Fig. 4a)

(Wang and LinHo 2002). This was expected because of the better capability of satellites to detect convective precipitation. The seasonal patterns observed in SAU are consistent with those reported by Ebert et al. (2007) and are attributed to the capabilities of reanalyses to capture synoptic precipitation (Fig. 4c). Reanalysis ETS in NAU outperformed satellite on an annual basis, and surprisingly, ERA-Interim was better than satellite precipitation during DJF and JJA. Ebert et al. (2007) attributed the better performance of reanalysis in NAU during JJA to remnant frontal systems brought in from midlatitudes or orographic lifting of moist ocean air during this season. Additional cause for better reanalysis performance in JJA and DJF may be due to the many grid cells in NAU close to SAU (Fig. 3a). Similar results

TABLE 1. Performance ranking of detection and accuracy metrics of precipitation products aggregated over the geographical domain, including means and standard deviations (in parentheses) of ETS, POD, FAR, FB, r , RMSD, and MPDR. Refer to section 2 for the definition of the metrics.

		Rank					
	1	2	3	4	5	6	7
ETS	Ensemble	JRA-25	Era-Interim	TRMM 3B42V6	NCEP-DOE	CMORPH	PERSIANN
	0.30 (± 0.05)	0.29 (± 0.05)	0.29 (± 0.06)	0.27 (± 0.05)	0.24 (± 0.04)	0.23 (± 0.06)	0.13 (± 0.04)
POD	Ensemble	JRA-25	ERA-Interim	NCEP-DOE	TRMM 3B42V6	CMORPH	PERSIANN
	0.62 (± 0.09)	0.62 (± 0.09)	0.62 (± 0.08)	0.55 (± 0.07)	0.52 (± 0.07)	0.49 (± 0.08)	0.36 (± 0.07)
FAR	ERA-Interim	Ensemble	TRMM 3B42V6	JRA-25	NCEP-DOE	CMORPH	PERSIANN
	0.40 (± 0.06)	0.41 (± 0.07)	0.41 (± 0.09)	0.42 (± 0.06)	0.42 (± 0.07)	0.42 (± 0.10)	0.54 (± 0.09)
FB	Ensemble	TRMM 3B42V6	JRA-25	NCEP-DOE	PERSIANN	CMORPH	ERA-Interim
	0.09 (± 0.46)	0.09 (± 0.11)	0.13 (± 0.14)	0.13 (± 0.17)	0.22 (± 0.23)	0.23 (± 0.21)	0.25 (± 0.33)
r	Ensemble	JRA-25	ERA-Interim	TRMM 3B42V6	CMORPH	PERSIANN	NCEP-DOE
	0.46 (± 0.09)	0.43 (± 0.08)	0.42 (± 0.08)	0.41 (± 0.09)	0.35 (± 0.10)	0.22 (± 0.23)	0.33 (± 0.17)
RMSD	Ensemble	TRMM 3B42V6	ERA-Interim	JRA-25	CMORPH	PERSIANN	NCEP-DOE
	6.43 (± 1.3)	6.56 (± 1.6)	6.60 (± 1.6)	6.62 (± 0.08)	6.68 (± 0.10)	7.54 (± 1.6)	10.7 (± 2.5)
MPDR	CMORPH	ERA-Interim	Ensemble	JRA-25	PERSIANN	TRMM 3B42V6	NCEP-DOE
	-1.3 (± 12.2)	-3.0 (± 8.10)	-4.9 (± 7.98)	5.33 (± 10.90)	-7.80 (± 11.8)	10.2 (± 6.97)	42.8 (± 13.3)

to those of ETS were observed for r , where the ensemble showed an equal or superior performance. Its ETS values were between both types of products because of the lower POD of satellite products especially during winter months (Tian et al. 2009). RMSD was similar for all products, with the exception of NCEP-DOE, which had a higher RMSD (particularly in NAU and SEA). NCEP-DOE also had higher positive MPDR than the other products. Large positive precipitation biases in the tropics have been reported for NCEP-DOE in other studies as well (Fekete et al. 2004; Bosilovich et al. 2008; Getirana et al. 2011). Surprisingly, although gauge-scaled, TRMM had systematically higher MPDR values in all geographical domains, and its RMSD was comparable to that of other satellite products. It has been argued that monthly scaling can propagate errors over space and time and that these may be reflected in RMSD (Gao and Liu 2012). In addition, a climatological undercatch correction is applied to TRMM (Huffman et al. 2007; Su et al. 2008), which is not present in precipitation analysis used for evaluation herein. Satellite product precipitation under or overestimation appeared to be location dependent (e.g., Ebert et al. 2007; Nesbitt et al. 2008; Romilly and Gebremichael 2011; Vernimmen et al. 2012), even for the gauge-scaled TRMM 3B42V6 product (Nair et al. 2009; Stampoulis and Anagnostou; 2012). Demaria et al. (2011) found that there was no clear gain of TRMM 3B42V6 over satellite products that are not bias corrected for precipitation exceeding 30 mm day⁻¹. They also showed that TRMM 3B42V6 would not necessarily improve estimates in areas with sparse gauges or if scaling introduces gauge data noise. Furthermore, because of data provider policies, it is not possible to know if some of the gauges used to calibrate

TRMM 3B42V6 are also part of the analysis data used here (Scheel et al. 2011).

Over all months combined and over the whole geographical domain, reanalysis outperformed satellite data on detection metrics and agreement metrics. Our results did, however, confirm the strength of satellite data in detecting and estimating convective precipitation.

By combining reanalyses and satellite products in an ensemble, known strengths of both retrieval systems resulted in a reduction of system-specific and random errors (e.g., Bosilovich et al. 2009). Issues associated with simple averaging of the products, such as a large bias in precipitation area and a corresponding reduction in mean and maximum precipitation depth (Ebert 2001), were addressed using a procedure that adjusts the probability distribution of the ensemble to the observed precipitation depth frequency. Although the depth-frequency-adjusted ensemble is not fully independent of the evaluation data, our results provide strong evidence that the inclusion of gauge information is valuable by adjusting both high and low precipitation depths. The dependence was limited, since the adjustment was performed over the whole geographical domain rather than by region or even by grid cell. An adjustment by subdomain or climate type could well improve estimates even further.

Acknowledgments. The authors gratefully acknowledge funding from the National Water Commission and Microsoft Research. We are also grateful for the assistance and/or correspondence to the providers of satellite and reanalyses data. Tim Raupach from CSIRO Land and Water and Beth Ebert from the Bureau of Meteorology are also thanked for reviewing the manuscript and for

providing valuable comments and suggestions. Tim also provided formatted Australian gauge and satellite data later used in this study.

REFERENCES

- Adler, R. F., G. Gu, and G. J. Huffman, 2012: Estimating Climatological bias errors for the Global Precipitation Climatology Project (GPCP). *J. Appl. Meteor. Climatol.*, **51**, 84–99.
- Behrangi, A., B. Khakbaz, T. C. Jaw, A. AghaKouchak, K. Hsu, and S. Sorooshian, 2011: Hydrologic evaluation of satellite precipitation products over a mid-size basin. *J. Hydrol.*, **397**, 225–237.
- Betts, A. K., M. Zhao, P. A. Dirmeyer, and A. C. M. Beljaars, 2006: Comparison of ERA40 and NCEP/DOE near-surface data sets with other ISLSCP-II data sets. *J. Geophys. Res.*, **111**, D22S04, doi:10.1029/2006JD007174.
- , M. Kohler, and Y. C. Zhang, 2009: Comparison of river basin hydrometeorology in ERA-Interim and ERA-40 reanalyses with observations. *J. Geophys. Res.*, **114**, D02101, doi:10.1029/2008JD010761.
- Bosilovich, M. G., J. Y. Chen, F. R. Robertson, and R. F. Adler, 2008: Evaluation of global precipitation in reanalyses. *J. Appl. Meteor. Climatol.*, **47**, 2279–2299.
- , D. Mocko, J. O. Roads, and A. Ruane, 2009: A multimodel analysis for the Coordinated Enhanced Observing Period (CEOP). *J. Hydrometeorol.*, **10**, 912–934.
- Dedong, L., Y. Zhongbo, H. Zhenchun, Y. Chuanguo, and J. Qin, 2007: Groundwater simulation in the Yangtze River basin with a coupled climate-hydrologic model. *J. China Univ. Geosci.*, **18**, 155–157.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597.
- Demaria, E. M. C., D. A. Rodriguez, E. E. Ebert, P. Salio, F. Su, and J. B. Valdes, 2011: Evaluation of mesoscale convective systems in South America using multiple satellite products and an object-based approach. *J. Geophys. Res.*, **116**, D08103, doi:10.1029/2010JD015157.
- Ebert, E. E., 2001: Ability of a poor man's ensemble to predict the probability and distribution of precipitation. *Mon. Wea. Rev.*, **129**, 2461–2480.
- , J. E. Janowiak, and C. Kidd, 2007: Comparison of near-real-time precipitation estimates from satellite observations and numerical models. *Bull. Amer. Meteor. Soc.*, **88**, 1–47.
- Fekete, B. M., C. J. Vorosmarty, J. O. Roads, and C. J. Willmott, 2004: Uncertainties in precipitation and their impacts on runoff estimates. *J. Climate*, **17**, 294–304.
- Fernandes, K., R. Fu, and A. K. Betts, 2008: How well does the ERA40 surface water budget compare to observations in the Amazon River basin? *J. Geophys. Res.*, **113**, D11117, doi:10.1029/2007JD009220.
- Gao, Y. C., and M. F. Liu, 2012: Evaluation of high-resolution satellite precipitation products using rain gauge observations over Tibetan Plateau. *Hydrol. Earth Syst. Sci. Discuss.*, **9**, 9503–9532.
- Getirana, A. C. V., J. C. V. Espinoza, J. Ronchail, and O. C. Rotunno Filho, 2011: Assessment of different precipitation data-sets and their impacts on the water balance of the Negro River basin. *J. Hydrol.*, **404**, 304–322.
- Gottschalk, J., J. Meng, M. Rodell, and P. Houser, 2005: Analysis of multiple precipitation products and preliminary assessment of their impact on global land data assimilation system land surface states. *J. Hydrometeorol.*, **6**, 573–598.
- Huffman, G. J., and Coauthors, 2007: The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.*, **8**, 38–55.
- Jeffrey, S. J., J. O. Carter, K. B. Moodie, and A. R. Beswick, 2001: Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Environ. Model. Software*, **16**, 309–330.
- Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. P. Xie, 2004: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeorol.*, **5**, 487–503.
- Kanamitsu, M., W. Ebisuzaki, J. Woollen, S. K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP-DOE AMIP-II reanalysis (R-2). *Bull. Amer. Meteor. Soc.*, **83**, 1631–1643.
- Khan, S. I., and Coauthors, 2012: Microwave satellite data for hydrologic modeling in ungauged basins. *IEEE Geosci. Remote Sens. Lett.*, **9**, 663–667.
- Li, J., X. Gao, and S. Sorooshian, 2008: Model performance of downscaling 1999–2004 hydrometeorological fields to the upper Rio Grande Basin using different forcing datasets. *J. Hydrometeorol.*, **9**, 677–694.
- Miguez-Macho, G., and Y. Fan, 2012: The role of groundwater in the Amazon water cycle: 1. Influence on seasonal streamflow, flooding and wetlands. *J. Geophys. Res.*, **117**, D15113, doi:10.1029/2012JD017539.
- Nair, S., G. Srinivasan, and R. Nemani, 2009: Evaluation of multi-satellite TRMM derived rainfall estimates over a western state of India. *J. Meteor. Soc. Japan*, **87**, 927–939.
- Nesbitt, S. W., D. J. Gochis, and T. J. Lang, 2008: The diurnal cycle of clouds and precipitation along the Sierra Madre Occidental observed during NAME-2004: Implications for warm season precipitation estimation in complex terrain. *J. Hydrometeorol.*, **9**, 728–743.
- Onogi, K., and Coauthors, 2007: The JRA-25 reanalysis. *J. Meteor. Soc. Japan*, **85**, 369–432.
- Pan, M., H. B. Li, and E. Wood, 2010: Assessing the skill of satellite-based precipitation estimates in hydrologic applications. *Water Resour. Res.*, **46**, W09535.
- Romilly, T. G., and M. Gebremichael, 2011: Evaluation of satellite rainfall estimates over Ethiopian river basins. *Hydrol. Earth Syst. Sci.*, **15**, 1505–1514.
- Ruane, A. C., and J. O. Roads, 2007: 6-hour to 1-year variance of five global precipitation sets. *Earth Interact.*, **11**. [Available online at <http://EarthInteractions.org>.]
- Sapiano, M. R. P., and P. A. Arkin, 2009: An intercomparison and validation of high-resolution satellite precipitation estimates with 3-hourly gauge data. *J. Hydrometeorol.*, **10**, 149–166.
- Scheel, M. L. M., M. Rohrer, C. Huggel, D. S. Villar, E. Silvestre, and G. J. Huffman, 2011: Evaluation of TRMM Multi-satellite Precipitation Analysis (TMPA) performance in the central Andes region and its dependency on spatial and temporal resolution. *Hydrol. Earth Syst. Sci.*, **15**, 8545–8586.
- Shrestha, M. S., G. A. Artan, S. R. Bajracharya, and R. R. Sharma, 2008: Using satellite-based rainfall estimates for streamflow modelling: Bagmati Basin. *J. Flood Risk Manage.*, **1**, 89–99.
- Sorooshian, S., K. L. Hsu, X. Gao, H. V. Gupta, B. Imam, and D. Braithwaite, 2000: Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bull. Amer. Meteor. Soc.*, **81**, 2035–2046.
- Stampoulis, D., and E. N. Anagnostou, 2012: Evaluation of global satellite rainfall products over continental Europe. *J. Hydrometeorol.*, **13**, 588–603.

- Su, F., Y. Hong, and D. P. Lettenmaier, 2008: Evaluation of TRMM Multisatellite Precipitation Analysis (TMPA) and its utility in hydrologic prediction in the La Plata Basin. *J. Hydrometeorol.*, **9**, 622–640.
- Tian, Y. D., and C. D. Peters-Lidard, 2010: A global map of uncertainties in satellite-based precipitation measurements. *Geophys. Res. Lett.*, **37**, L24407, doi:10.1029/2010GL046008.
- , and Coauthors, 2009: Component analysis of errors in satellite-based precipitation estimates. *J. Geophys. Res.*, **114**, D24101, doi:10.1029/2009JD011949.
- Uppala, S., A. Simmons, D. Dee, P. Kallberg, and J. N. Thepaut, 2007: Atmospheric reanalyses and climate variations. *Climate Variability and Extremes During the Past 100 Years*, S. Brönnimann et al., Eds., Advances in Global Change Research, Vol. 33, Springer, 103–117.
- Van Dijk, A. I. J. M., and L. J. Renzullo, 2011: Water resource monitoring systems and the role of satellite observations. *Hydrol. Earth Syst. Sci.*, **15**, 39–55.
- Vernimmen, R. R. E., A. Hooijer, E. Mamenun, Aldrian, and A. I. J. M. van Dijk, 2012: Evaluation and bias correction of satellite rainfall data for drought monitoring in Indonesia. *Hydrol. Earth Syst. Sci.*, **16**, 133–146.
- Vila, D., R. Ferraro, and H. Semunegus, 2010: Improved global rainfall retrieval using the Special Sensor Microwave Imager (SSM/I). *J. Appl. Meteor. Climatol.*, **49**, 1032–1043.
- Voisin, N., A. W. Wood, and D. P. Lettenmaier, 2008: Evaluation of precipitation products for global hydrological prediction. *J. Hydrometeorol.*, **9**, 388–407.
- Wang, B., and LinHo, 2002: Rainy season of the Asian–Pacific summer monsoon. *J. Climate*, **15**, 386–398.
- Yatagai, A., K. Kamiguchi, O. Arakawa, A. Hamada, N. Yasutomi, and A. Kitoh, 2012: APHRODITE: Constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges. *Bull. Amer. Meteor. Soc.*, **93**, 1401–1415.
- Yong, B., Y. Hong, L.-L. Ren, J. J. Gourley, G. J. Huffman, X. Chen, W. Wang, and S. I. Khan, 2012: Assessment of evolving TRMM-based multisatellite real-time precipitation estimation methods and their impacts on hydrologic prediction in a high latitude basin. *J. Geophys. Res.*, **117**, D09108, doi:10.1029/2011JD017069.